PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G10L 5/06, H04M 1/64

A1

(11) International Publication Number: WO 99/56275

(43) International Publication Date: 4 November 1999 (04.11.99)

(21) International Application Number: PCT/US99/06600

(22) International Filing Date: 25 March 1999 (25.03.99)

(30) Priority Data:
09/067,779 27 April 1998 (27.04.98) US

(71) Applicant: MOTOROLA INC. [US/US]; 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

(72) Inventors: CHENG, Yan, Ming; 2129 Kensington Drive, Schaumburg, IL 60194 (US). CHANG, Lu; 7747 Villa Nova Drive, Boca Raton, FL 33433 (US). FINLON, Kenneth, D.; 15653 Bent Creek Road, Wellington, FL 33414 (US). HERKERT, Arthur, G.; 54 Cedar Circle, Lantana, FL 33462 (US).

(74) Agents: NICHOLS, Daniel, K. et al.; Motorola Inc., Intellectual Property Dept., 1500 Gateway Boulevard, Boynton Beach, FL 33426-8292 (US).

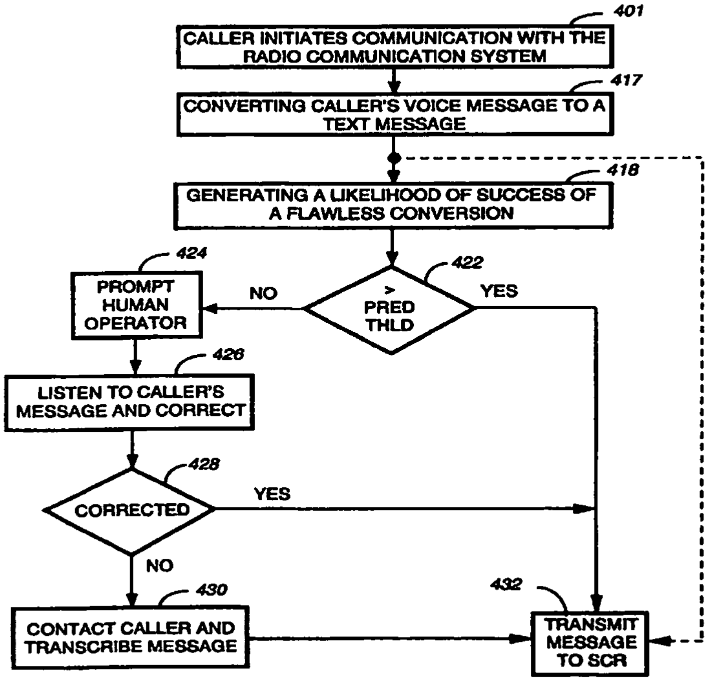
(81) Designated States: BR, CA, CN, JP, MX, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
*With international search report.*

## (54) Title: RELIABLE TEXT CONVERSION OF VOICE IN A RADIO COMMUNICATION SYSTEM AND METHOD

## (57) Abstract

A radio communication system includes a voice recognition system (218), a transmitter (202), and a processing system (210). The transmitter (202) is used for transmitting messages to a plurality of SCRs (selective call radios) (122). The processing system (210) is adapted to cause the voice recognition system (218) to convert a voice signal representative of a voice message originated by a caller to a text message (401, 417), wherein the text message is intended for a SCR (122), and to cause the transmitter (202) to transmit the text message (401, 417) to the SCR (122). An embodiment of the voice recognition system (218) may also generate a likelihood of success of flawless conversion (418), and the processing system will transmit the text message or prompt a human operator (424) to generate a corrected text message based on an accuracy threshold (422).


401
CALLER INITIATES COMMUNICATION WITH THE RADIO COMMUNICATION SYSTEM
417
CONVERTING CALLER'S VOICE MESSAGE TO A TEXT MESSAGE
418
GENERATING A LIKELIHOOD OF SUCCESS OF A FLAWLESS CONVERSION
424
PROMPT HUMAN OPERATOR
NO
422
> PRED THLD
YES
426
LISTEN TO CALLER'S MESSAGE AND CORRECT
428
CORRECTED
YES
NO
430
CONTACT CALLER AND TRANSCRIBE MESSAGE
432
TRANSMIT MESSAGE TO SCR

400

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AL | Albania |
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| AZ | Azerbaijan |
| BA | Bosnia and Herzegovina |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CU | Cuba |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GH | Ghana |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IL | Israel |
| IS | Iceland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgyzstan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakstan |
| LC | Saint Lucia |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LS | Lesotho |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| MK | The former Yugoslav Republic of Macedonia |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TM | Turkmenistan |
| TR | Turkey |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |
| YU | Yugoslavia |
| ZW | Zimbabwe |

# RELIABLE TEXT CONVERSION OF VOICE IN A RADIO COMMUNICATION SYSTEM AND METHOD

## Related Invention

The present invention is related to the following invention which is assigned to the same assignee as the present invention:

U.S. Application No. 09/050,184 filed March 30, 1998 by Andric et al., entitled "Voice Recognition System in a Radio Communication System and Method Therefor."

## Field of the Invention

This invention relates in general to radio communication systems, and particularly, to reliable conversion of voice in a radio communication system.

## Background of the Invention

Presently, voice recognition systems are becoming popular with consumers of conventional computers due to the availability of continuous speech recognition systems. These applications are generally tailored for speaker-dependent voice recognition. That is, to provide a high degree of accuracy in the conversion of voice to a textual message, the continuous speech recognition system must be trained by a particular speaker's voice. This is generally performed by having the speaker read a canned message of several paragraphs, which is then recorded and analyzed by the speech recognition system to develop a set of statistical models of the speaker's voice. These models are later used by the continuous speech recognition system to convert the speaker's voice signals into a textual message.

Although the present approach provides a relatively high degree of accuracy in the process of converting voice to a textual message, a need for higher degrees of accuracy approaching a flawless conversion is desirable. Present continuous speech recognition systems suffer several disadvantages in reaching a flawless conversion standard. For example, present algorithms rely heavily on the spectral envelope features of the analyzed speech signals to generate a textual message equivalent. This is disadvantageous because such a method fails to account for other features in speech such as the shape of the speech spectrum, which may be helpful in improving the accuracy of voice conversion. Additionally, present algorithms are not well adapted to recognize speech at a high degree of accuracy from speakers who have not trained the system with their particular voice characteristics.

Because of the foregoing limitations in prior art voice recognition systems, service providers of radio communication systems have opted to utilize human operators to transcribe voice messages to text messages from callers who intend to send messages to one or more SCRs (selective call radios) of the radio communication system. Service providers are generally hesitant in using a completely automated voice recognition system, because present voice recognition systems cannot guarantee flawless conversion of voice messages to text messages. The use of human operators, however, is expensive, especially for radio communication systems that operate 24 hours a day, every day of the year. Consequently, a need exists for automating the conversion of voice messages to text messages in a radio communication system to the extent that reliance on human operators to perform this conversion is either eliminated or substantially reduced.

Accordingly, what is needed is an apparatus and method for reliable conversion of voice in a radio communication system that satisfies present needs, and overcomes the foregoing disadvantages in the prior art.

## Brief Description of the Drawings

The present invention is pointed out with particularity in the appended claims. However, other features of the invention will become more apparent and best understood by referring to the following detailed description in conjunction with the accompanying drawings in which:

FIG. 1 is an electrical block diagram of a radio communication system according to the present invention;

FIGs. 2 and 3 are electrical block diagrams of the fixed and portable portions of the radio communication system according to the present invention;

FIG. 4 depicts the use of a human operator in the radio communication system according to the present invention;

FIGs. 5-6 show flowcharts summarizing the operation of the radio communication system according to the present invention; and

FIGs. 7-10 show graphs representative of the transformations made to voice signals generated by a caller according to the present invention.

## Description of the Preferred Embodiment

FIG. 1 is an electrical block diagram of a radio communication system comprising a fixed portion 102 and a portable portion 104. The fixed portion 102 includes a controller 112 for controlling operation of a plurality of base stations 116 by way of conventional communication links 114, such as microwave links. The portable portion 104 includes a plurality of SCR's (selective call radios) 122 for receiving messages from the base stations 116 under the control of the controller 112. It will be appreciated that, alternatively, the radio communication system may be modified to support two-way communication between the SCR's 122 and the base stations 116. This modification may be achieved by the use of radio transceivers at both the SCR's 122 and the base stations 116.

Turning to the operation of the controller 112, we find that the controller 112 receives messages from callers utilizing a conventional telephone 124 for communicating with a conventional PSTN (public switch telephone network) 110. The PSTN 110 then relays messages to the controller 112 through a conventional telephone line 101 coupled to the controller 112. Upon receiving messages from the PSTN 110, the controller 112 processes the messages, and delivers them to the base stations 116 for transmission to designated SCR's 122. It will be appreciated that, alternatively, the telephone 124 may be directly coupled to the controller 112 by way of a conventional telephone line 103.

FIGs. 2 and 3 are electrical block diagrams of the fixed and portable portions 102, 104 of the radio communication system according to the present invention. The electrical block diagram of the fixed portion 102 includes the elements of the controller 112 and the base stations 116. The controller 112 comprises a conventional processing system 210 for controlling operation of the base stations 116, a voice recognition system 218, and a transmitter interface 204 for communicating messages to the base stations 116. The voice recognition system 218 receives voice messages from the PSTN 110, and/or from a direct telephone connection 103, and converts the voice messages to equivalent text messages. The processing system 210 includes conventional hardware such as a computer system 212 (with built-in random access memory (RAM)--not shown in FIG. 2) and mass media 214 (e.g., a conventional hard disk) to perform the programmed operations of the controller 112. The base stations 116 comprise a conventional RF transmitter 202 coupled to an antenna 201 for transmitting the messages received from the controller 112.

A detailed discussion of the SCR 122 will be postponed until after the fixed portion 102 has been discussed. To start this discussion, the reader is directed to FIGs. 5-6, which show flowcharts 400, 417 summarizing the operation of the radio communication system according to the present invention. The flowchart 400 depicts programmed instructions of the controller 112 which are initially stored in the mass media 214 and are then operated from the RAM included in the computer system 212.

The flowchart 400 begins with step 401 where a caller initiates communication with the radio communication system intending to send a message to a selected SCR 122. As noted earlier, this communication may originate from the PSTN 110 or a direct telephone connection 103 with the controller 112. In step 417, the caller's voice signal is converted to a textual equivalent of speech. After the conversion step 417, in a first embodiment, the text message is directly transmitted the selected SCR 122 in step 432. No further processing is required of the processing system 210. In an alternative embodiment, after the conversion step 417, the processing system 210 proceeds to step 418 where the voice recognition system 218 generates a likelihood of success (e.g., between 0% and 100%) that the voice signal has been flawlessly converted to a text message. In step 422, the processing system 210 compares the likelihood of success to a predetermined threshold.

The predetermined threshold is, for example, selected by the service provider of the radio communication system based on a minimum acceptable level of accuracy desired (e.g., confidence level below 90% is unacceptable) from the conversion step 417. In the event that likelihood of success generated in step 418 is below the predetermined threshold chosen, then the processing system 210 proceeds to step 424; otherwise, the processing system 210 proceeds to step 432 where the text message is transmitted to the targeted SCR 122. Turning to step 424, the processing system 210 prompts a human operator of the radio communication system to listen to an audible representation of the voice signal generated by the caller in step 401, and to generate a corrected text message in step 426.

Step 426 may be accomplished, for example, by having the human operator sit at a computer terminal (see FIG. 4) coupled to the radio communication system, listening to the audible representation of the voice signal and transcribing at the computer terminal the caller's voice message in total. Once this has been completed, the human operator presents the corrected text message to the radio communication by prompting the radio communication system to accept the corrected text message. Presentation of the corrected message may be accomplished by depressing one or more predetermined keys on the computer terminal (e.g., CTRL T representative of a command to transmit the text message to the SCR 122). It will be appreciated that alternative conventional methods may exist for delivering the corrected text message to the radio communication system, and that any of these methods would be considered by one of ordinary skill in the art to be within the scope of the present invention.

It will be further appreciated that, alternatively, the human operator may listen audibly to the caller's voice message, and view contemporaneously the text message generated by the voice recognition system 218 in step 417 on a monitor of the computer terminal which the human operator is operating from. In doing so, the human operator may find that the text message was converted flawlessly, and no correction is necessary. It is worth noting that a likelihood of success below the predetermined threshold (e.g., 90%) does not necessarily mean that the conversion of the caller's voice message was flawed. For this reason, the human operator may find that a correction is not necessary after listening to the audible representation of the voice signal. Similarly, the human operator may find that the text message generated by the voice recognition system 218 in step 417 only has a few errors. In that case, the human operator would correct these flaws rather than transcribe the entire message.

Lastly, in the event that the human operator is unable to interpret the caller's voice message in step 426, the above embodiments describing step 426 are modified such that the controller 112 places the caller on hold while still communicating through the PSTN 110 or the direct telephone line 103 with the radio communication system. Once the human operator finds that the audible representation of the voice signal is incomprehensible in step 426, the human operator proceeds to contact the caller in step 428 and requests a repetition of the voice message in step 430. The human operator then transcribes the repeated voice message into a corrected text message.

Upon completion of any of the foregoing embodiments depicted by steps 417-430, the processing system 210 proceeds to step 432 whereby it causes a selected base station 116 to transmit the corrected text message to the SCR 122.

A prominent feature of the present invention which substantially reduces the use of a human operator as depicted in steps 424-430 of FIG.

5 is included in the voice recognition system 218. Although the present invention is not limited in scope to a single type of voice recognition system, the flowchart of FIG. 6 illustrates a preferred embodiment of the voice recognition system 218. This embodiment provides a high degree of first-time success in converting a caller's voice message to a textual message flawlessly, thereby limiting the frequency for which steps 424-430 are invoked.

The process of converting voice to a textual message begins with step 402 where a voice signal originated by a caller in step 401 is sampled. An illustration of a voice signal is shown in FIG. 7. In step 403 the processing system 210 is programmed to apply a Fourier transform to a plurality of frame intervals of the sampled voice signal (e.g., 10-25 ms) to generate spectral data having a spectral envelope for each of the plurality of frame intervals. The Fourier transform applied in this step is preferably a fast Fourier transform. The spectral signal over a frame interval is shown in FIG. 8. Assuming the input speech signal is represented by $x_n$, the following equation describes the result of step 403:

$$P_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi nk}{N}},$$

where $0 \le k \le N - 1$.

In step 404, for each of the plurality of frame intervals, the spectral data is subdivided into a plurality of bands, each of the plurality of bands having a predetermined bandwidth (e.g., 400 Hz). It will be appreciated that, alternatively, each band may be of variable bandwidth. In step 406, the processing system 210 determines an average magnitude of the spectral data for each of the plurality of bands. Then in step 407 a logarithmic function is applied to the average magnitude to generate a converted average magnitude. In step 408, the converted average magnitude is then decorrelated (preferably with a discrete cosine transform) to generate spectral envelope features.

The controller 112 then proceeds to step 409 to filter out the spectral envelope from the spectral data of each of the plurality of frame intervals to generate filtered spectral data for each of the plurality of frame intervals. This step preferably comprises the steps of averaging the spectral data of each of the plurality of frame intervals to generate a spectral envelope estimate, and subtracting the spectral envelope estimate from the spectral data. These steps are substantially represented by the function,

$$P'_k = f(i) * P_k \text{ , wherein } f(i) = \begin{cases} 1 & 0 \le i < L \\ -1 & -L < i < 0 \end{cases}.$$

The function $f(i)$ is a 1-D Haar function well known in the art, and $P'_k$ is the convolution of the Haar function with the original spectral data $P_k$. The result of filtering the spectral data is shown in FIG. 9.

Next, in step 410, a fast Fourier transform is applied to the filtered spectral data for each of the plurality of bands to generate an autocorrelation function for each of the plurality of bands. If there is a strong harmonic structure in the original spectral data, the autocorrelation function for each of the plurality of bands will have a high peak value around the value of its pitch period. For this reason, each autocorrelation function is preferably normalized by its corresponding spectral band energy. In step 412, the controller 112 proceeds to measure a value of the magnitude of the autocorrelation function for each of the plurality of bands. The value of the magnitude of the autocorrelation function is defined as a measure of a degree of voiceness for each of the plurality of bands.

There are two embodiments for measuring a value of the magnitude of the autocorrelation function. In a first embodiment, the value of the magnitude of the autocorrelation function corresponds to a peak magnitude of the autocorrelation function. Alternatively, in a second embodiment, for each of the plurality of frame intervals, the value of the magnitude of the autocorrelation function for each of the plurality of bands is determined by: (1) summing the autocorrelation function of each of the plurality of bands to generate a composite autocorrelation function, (2) determining a peak magnitude of the composite autocorrelation function, (3) determining from the peak magnitude a corresponding frequency mark, and (4) utilizing the corresponding frequency mark to determine a corresponding magnitude value for each of the plurality of bands.

The second embodiment is illustrated in FIG. 10. Graphs (a)-(d) represent the autocorrelation function for each of bands 1-4. Graph (e) is the composite autocorrelation function as a result of summing the autocorrelation functions of bands 1-4. From the composite autocorrelation function a peak magnitude, and a corresponding frequency mark is determined. The corresponding frequency mark is then used to determine a corresponding magnitude value for each of the plurality of bands as shown in graphs (a)-(d).

As noted earlier, the value of the magnitude of the autocorrelation function is a measure of the degree of voiceness for each of the plurality of bands. After determining the degree of voiceness for each of the plurality of bands by either of the foregoing embodiments, in step 414, the spectral envelope features determined in step 408 and the degree of voiceness just discussed is applied to a corresponding plurality of phoneme models. Phoneme models are known in the art as models of speech determined from statistical modeling of human speech. In the art, phoneme models are also commonly referred to as Hidden Markov Models. A phoneme represents the smallest quantum of sound used by a speaker for constructing a word. For example, the word “is” may be decomposed into two phoneme sounds: “ih” and “z.” Since individuals of differing cultures may speak with differing dialects, the word “is” may have more than one set of phoneme models to represent mismatched populations. For example, there may be individuals who end the word “is” with a “s” sound, i.e., “ih” and “s.”

As a preferred embodiment, the phoneme models are determined over a large population of samples of human speech, which accounts for varying pronunciations based on varying speech dialectics. Deriving phoneme models from a large population allows for the present invention to operate as a speaker-independent voice recognition system. That is, the phoneme models are not dependent on a particular speaker’s voice. With speaker-independent descriptions built into a phoneme model library, the controller 112 of the radio communication system can convert the voice of any speaker nation-wide without prior training of the caller's voice to a textual message. It will be appreciated, however, that the present invention may be altered so that a phoneme library may be constructed from training provided by one or more specific speakers, thereby forming a speaker-dependent phoneme library. Notwithstanding this alternative embodiment, the ensuing discussions will focus on a speaker-independent phoneme library.

Based on a speaker-independent phoneme library, the conversion of voice into a textual message, as indicated by step 416, is accomplished by comparing the spectral envelope features of the spectral data for each of the plurality of bands and the degree of voiceness for each of the plurality of bands with a library of speaker-independent phoneme models. From this comparison, a list of likely phonemes are identified, which are then compared to a dictionary of words (from, e.g., the English language) and their corresponding phonemes to derive a textual equivalent of speech from the processed voice signal of the caller. As part of the comparison processes for determining one or more likely phonemes, the following probability function is preferably used:

$$b_j(o_t) = \prod_{s=1}^{S} \left[ \sum_{m=1}^{M_s} c_{jsm} N\left(o_{st}; \mu_{jsm}, \Sigma_{jsm}\right) \right]^{\gamma_s},$$

wherein $M_s$ is the number of mixture components in stream $s$. The variable $S$ for the present invention is equal to 2, which represents the product of two probabilities: That is, one product represents the likelihood of a matched set of phoneme models based on the spectral envelope features of the spectral data per band, and another product represents the likelihood of a matched set of phoneme models based on the degree of voiceness per band. The variable $c_{jsm}$ is weighting factor, while the function $N$ is a multivariate Gaussian function, wherein the variable $o_{st}$ is input data vectors representative of the spectral envelope features and degree of voiceness for each of the plurality of bands, and wherein $u_{jsm}$ and $\Sigma_{jsm}$ are the mean and covariance vectors of each of the phoneme models in the phoneme library. Lastly, the variable $\gamma_s$ is used for providing differing weights to the spectral envelope features probability result versus the degree of voiceness probability result. For example, the spectral envelope features probability result may be given a weight of 1.00 while the degree of voiceness probability result may be given a weight of 1.20. Hence, more importance is given to the outcome derived from the use of degree of voiceness data rather than the spectral envelope features data. It will be appreciated that any weight may be given to either product depending on the application in which the present invention is utilized.

Each of the probability results ($b_j$) is then compared over a stream of the plurality of frames to determine an equivalent textual version of the caller's voice message. In the event the comparison process leads to one or more possible textual streams, the textual stream with the greatest likelihood of success is chosen according to a composite probability result for each branch. Once the textual result with the greatest likelihood of success has been chosen, the controller 112 proceeds to steps 418-426 of FIG. 5 as discussed earlier.

This article is useful for gaining further insight into the use of the foregoing equation (represented by $b_j$) to predict the likelihood of a stream of phonemes derived from a voice signal.

A detailed description of the foregoing equation (represented by $b_j$) to predict the likelihood of a stream of phonemes is more fully described in Steve Young, "The HTK Book," Entropic Cambridge Research Laboratory, Cambridge CB3 OAX, England, which is hereby incorporated herein by reference. Additionally, the reader is directed to the following introductory materials related to voice recognition systems, which are described in Joseph Picone, "Continuous Speech Recognition Using Hidden Markov Models," IEEE ASSP Magazine, July 1990, pp. 26-40, and Yves Normandin, "High-Performance Connected Digit Recognition Using Maximum Mutual Information Estimation," IEEE Transactions on Speech and Audio Processing, Vol. 2, No. 2, April 1994, respectively, which are hereby incorporated herein by reference.

The foregoing method and apparatus are substantially advantages over prior art systems. First, the use of a voice recognition system for converting voice messages to text messages substantially reduces the need for human operators for transcribing messages, thereby reducing cost. Second, although not necessarily required for the present invention, employing a preferred embodiment for the operation of the voice recognition system 218 as depicted by the flowchart of FIG. 5 adds further improvement to the present invention over the prior art. Particularly, the reader is reminded from the background of the invention that prior art systems have a limited success rate of converting voice messages to textual messages due to an emphasis put on deriving textual messages based on the spectral envelope features of the analyzed speech signal.

In contrast, the present invention takes advantage of analyzing the texture of the speech spectrum (described above as the degree of voiceness) along with the spectral envelope features of the speech signal. By utilizing both magnitude data of the spectral signal and degree of voiceness data for comparison to a phoneme library, the present invention provides a higher degree of accuracy for flawlessly converting speaker-dependent and speaker-independent voice signals to a text message.

Having summarized the fixed portion 102 of the radio communication system, the reader's attention is now directed to FIG. 3, which shows an electrical block diagram of the SCR 122 according to the present invention. As noted in step 432 of FIG. 5, the SCR 122 receives textual messages (in the form of, e.g., alpha-numeric messages) generated by a caller after having been processed by the fixed portion 102 of the radio communication as described by the flowcharts of FIGs. 5-6. The SCR 122 comprises a receiver 304 coupled to an antenna 302, a power switch 306, a processor 308, an alerting device 316, a display 318, and user controls 314. The receiver 304 and antenna 302 are conventional RF elements for receiving messages transmitted by the base stations 116. The power switch 306 is a conventional switch, such as a MOS (metal oxide semiconductor) switch for controlling power to the receiver 304 under the direction of the processor 308, thereby providing a battery saving function.

The processor 308 is used for controlling operation of the SCR 122. Generally, its primary function is to decode and process demodulated messages provided by the receiver 304, storing them and alerting a user of the received message. To perform this function, the processor 308 comprises a conventional microprocessor 312 coupled to a conventional memory 310 including nonvolatile and volatile memory portions, such as a ROM (read-only memory) and RAM (random-access memory). One of the uses of the memory 310 is for storing messages received from the base stations 116. Another use is for storing one or more selective call addresses utilized in identifying incoming messages belonging to the SCR 122.

Once a message has been decoded and stored in the memory 310, the processor 308 activates the alerting device 316 which generates a tactile and/or audible alert signal to the user. With the display 318, which is, for example, a conventional LCD (liquid crystal display) and conventional user controls 314, the user may process the received messages. The user controls 314 provide options such as reading, deleting, and locking of messages.

Although the invention has been described in terms of a preferred embodiment it will be obvious to those skilled in the art that many alterations and variations may be made without departing from the invention. Accordingly, it is intended that all such alterations and variations be considered as within the spirit and scope of the invention as defined by the appended claims.

What is claimed is:

## CLAIMS

1. In a radio communication system, a method comprising the steps of:

converting a voice signal representative of a voice message originated by a caller to a text message, wherein the text message is intended for a SCR (selective call radio);

generating a likelihood of success that the voice signal has been flawlessly converted to a text message;

comparing the likelihood of success to a predetermined threshold;

if the likelihood of success is below the predetermined threshold, prompting a human operator of the radio communication system to:

listen to an audible representation of the voice signal, and

generate a corrected text message; and

transmitting the corrected text message to the SCR.

2. The method as recited in claim 1, wherein the correcting step comprises the steps of:

placing the caller on hold while the human operator listens to the audible representation of the voice signal;

if the human operator cannot interpret the audible representation of the voice signal, prompting the human operator to:

contact the caller to repeat the voice message, and

transcribe the repeated voice message to the corrected text message; and

transmitting the corrected text message to the SCR.

3. The method as recited in claim 1, wherein the converting step comprises the steps of:

sampling a voice signal;

applying a Fourier transform to a plurality of frame intervals of the sampled voice signal to generate spectral data having a spectral envelope for each of the plurality of frame intervals;

subdividing the spectral data for each of the plurality of frame intervals into a plurality of bands;

filtering out the spectral envelope from the spectral data of each of the plurality of frame intervals to generate filtered spectral data for each of the plurality of frame intervals;

applying a Fourier transform to the filtered spectral data for each of the plurality of bands to generate an autocorrelation function for each of the plurality of bands;

measuring a value of the magnitude of the autocorrelation function for each of the plurality of bands, whereby the value is a measure of a degree of voiceness for each of the plurality of bands;

applying the degree of voiceness for each of the plurality of bands to a corresponding plurality of phoneme models; and

deriving a textual equivalent of speech from the voice signal by searching through a phoneme library according to predictions made by the corresponding plurality of phoneme models.

4. The method as recited in claim 3, further comprising the steps of:

determining an average magnitude for each of the plurality of bands;

applying a logarithmic function to the average magnitude to generate a converted average magnitude;

decorrelating the converted average magnitude to generate spectral envelope features; and

applying the spectral envelope features for each of the plurality of bands to the corresponding plurality of phoneme models.

5. The method as recited in claim 3, wherein the value of the magnitude of the autocorrelation function is a peak magnitude.

6. The method as recited in claim 3, wherein for each of the plurality of frame intervals, the value of the magnitude of the autocorrelation function for each of the plurality of bands is determined by:

summing the autocorrelation function of each of the plurality of bands to generate a composite autocorrelation function;

determining a peak magnitude of the composite autocorrelation function;

determining from the peak magnitude a corresponding frequency mark; and

utilizing the corresponding frequency mark to determine a corresponding magnitude value for each of the plurality of bands.

7. The method as recited in claim 3, further comprising the step of normalizing the autocorrelation function for each of the plurality of bands by its corresponding spectral band energy.

8. The method as recited in claim 3, wherein the Fourier transform comprises a fast Fourier transform.

9. The method as recited in claim 3, wherein the step of filtering out the spectral envelope comprises the steps of:

averaging the spectral data of each of the plurality of frame intervals to generate a spectral envelope estimate; and

subtracting the spectral envelope estimate from the spectral data of each of the plurality of frame intervals.

10. In a radio communication system, a method comprising the steps of:

converting a voice signal representative of a voice message originated by a caller to a text message, wherein the text message is intended for a SCR (selective call radio); and

transmitting the text message to the SCR.

11. The method as recited in claim 10, wherein the converting step comprises the steps of:

sampling a voice signal;

applying a Fourier transform to a plurality of frame intervals of the sampled voice signal to generate spectral data having a spectral envelope for each of the plurality of frame intervals;

subdividing the spectral data for each of the plurality of frame intervals into a plurality of bands;

filtering out the spectral envelope from the spectral data of each of the plurality of frame intervals to generate filtered spectral data for each of the plurality of frame intervals;

applying a Fourier transform to the filtered spectral data for each of the plurality of bands to generate an autocorrelation function for each of the plurality of bands;

measuring a value of the magnitude of the autocorrelation function for each of the plurality of bands, whereby the value is a measure of a degree of voiceness for each of the plurality of bands;

applying the degree of voiceness for each of the plurality of bands to a corresponding plurality of phoneme models; and

deriving a textual equivalent of speech from the voice signal by searching through a phoneme library according to predictions made by the corresponding plurality of phoneme models.

12. The method as recited in claim 11, further comprising the steps of:

determining an average magnitude for each of the plurality of bands;

applying a logarithmic function to the average magnitude to generate a converted average magnitude;

decorrelating the converted average magnitude to generate spectral envelope features; and

applying the spectral envelope features for each of the plurality of bands to the corresponding plurality of phoneme models.

13. The method as recited in claim 11, wherein the value of the magnitude of the autocorrelation function is a peak magnitude.

14. The method as recited in claim 11, wherein for each of the plurality of frame intervals, the value of the magnitude of the autocorrelation function for each of the plurality of bands is determined by:

summing the autocorrelation function of each of the plurality of bands to generate a composite autocorrelation function;

determining a peak magnitude of the composite autocorrelation function;

determining from the peak magnitude a corresponding frequency mark; and

utilizing the corresponding frequency mark to determine a corresponding magnitude value for each of the plurality of bands.

15. The method as recited in claim 11, further comprising the step of normalizing the autocorrelation function for each of the plurality of bands by its corresponding spectral band energy.

16. The method as recited in claim 11, wherein the Fourier transform comprises a fast Fourier transform.

17. The method as recited in claim 11, wherein the step of filtering out the spectral envelope comprises the steps of:

averaging the spectral data of each of the plurality of frame intervals to generate a spectral envelope estimate; and

subtracting the spectral envelope estimate from the spectral data of each of the plurality of frame intervals.

18. A radio communication system, comprising:

a voice recognition system for receiving caller initiated messages;

a transmitter for transmitting messages to a plurality of SCRs (selective call radios) of the radio communication system; and

a processing system coupled to the voice recognition system, and the transmitter, wherein the processing system is adapted to:

cause the voice recognition system to convert a voice signal representative of a voice message originated by a caller of the radio communication system to a text message, wherein the text message is intended for a SCR;

cause the voice recognition system to generate a likelihood of success that the voice signal has been flawlessly converted to a text message;

compare the likelihood of success to a predetermined threshold;

if the likelihood of success is below the predetermined threshold, prompting a human operator of the radio communication system to:

listen to an audible representation of the voice signal, and

generate a corrected text message; and

cause the transmitter to transmit the corrected text message to the SCR.

19. The radio communication system as recited in claim 18, wherein the correcting step the processing system is further adapted to:

place the caller on hold while the human operator listens to the audible representation of the voice signal;

if the human operator cannot interpret the audible representation of the voice signal, prompt the human operator to:

contact the caller to repeat the voice message, and

transcribe the repeated voice message to the corrected text message; and

cause the transmitter to transmit the corrected text message to the SCR.

20. The radio communication system as recited in claim 18, wherein the voice recognition system is adapted to:

sample a voice signal generated by a caller during a plurality of frame intervals, wherein the voice signal is representative of a message intended for a selective call radio;

apply a Fourier transform to a plurality of frame intervals of the sampled voice signal to generate spectral data having a spectral envelope for each of the plurality of frame intervals;

subdivide the spectral data for each of the plurality of frame intervals into a plurality of bands;

filter out the spectral envelope from the spectral data of each of the plurality of frame intervals to generate filtered spectral data for each of the plurality of frame intervals;

apply a Fourier transform to the filtered spectral data for each of the plurality of bands to generate an autocorrelation function for each of the plurality of bands;

measure a value of the magnitude of the autocorrelation function for each of the plurality of bands, whereby the value is a measure of a degree of voiceness for each of the plurality of bands;

apply the degree of voiceness for each of the plurality of bands to a corresponding plurality of phoneme models;

derive a textual equivalent of speech from the voice signal by searching through a phoneme library according to predictions made by the corresponding plurality of phoneme models; and

cause the transmitter to transmit the textual equivalent of speech to the selective call radio, wherein the textual equivalent of speech is representative of the message initiated by the caller.

21. A radio communication system, comprising:

a voice recognition system for receiving caller initiated messages;

a transmitter for transmitting messages to a plurality of SCRs (selective call radios) of the radio communication system; and

a processing system coupled to the voice recognition system, and the transmitter, wherein the processing system is adapted to:

cause the voice recognition system to convert a voice signal representative of a voice message originated by a caller of the radio communication system to a text message, wherein the text message is intended for a SCR; and

cause the transmitter to transmit the text message to the SCR.

22. The radio communication system as recited in claim 21, wherein the voice recognition system is adapted to:

sample a voice signal generated by a caller during a plurality of frame intervals, wherein the voice signal is representative of a message intended for a selective call radio;

apply a Fourier transform to a plurality of frame intervals of the sampled voice signal to generate spectral data having a spectral envelope for each of the plurality of frame intervals;

subdivide the spectral data for each of the plurality of frame intervals into a plurality of bands;

filter out the spectral envelope from the spectral data of each of the plurality of frame intervals to generate filtered spectral data for each of the plurality of frame intervals;

apply a Fourier transform to the filtered spectral data for each of the plurality of bands to generate an autocorrelation function for each of the plurality of bands;

measure a value of the magnitude of the autocorrelation function for each of the plurality of bands, whereby the value is a measure of a degree of voiceness for each of the plurality of bands;

apply the degree of voiceness for each of the plurality of bands to a corresponding plurality of phoneme models;

derive a textual equivalent of speech from the voice signal by searching through a phoneme library according to predictions made by the corresponding plurality of phoneme models; and

cause the transmitter to transmit the textual equivalent of speech to the selective call radio, wherein the textual equivalent of speech is representative of the message initiated by the caller.

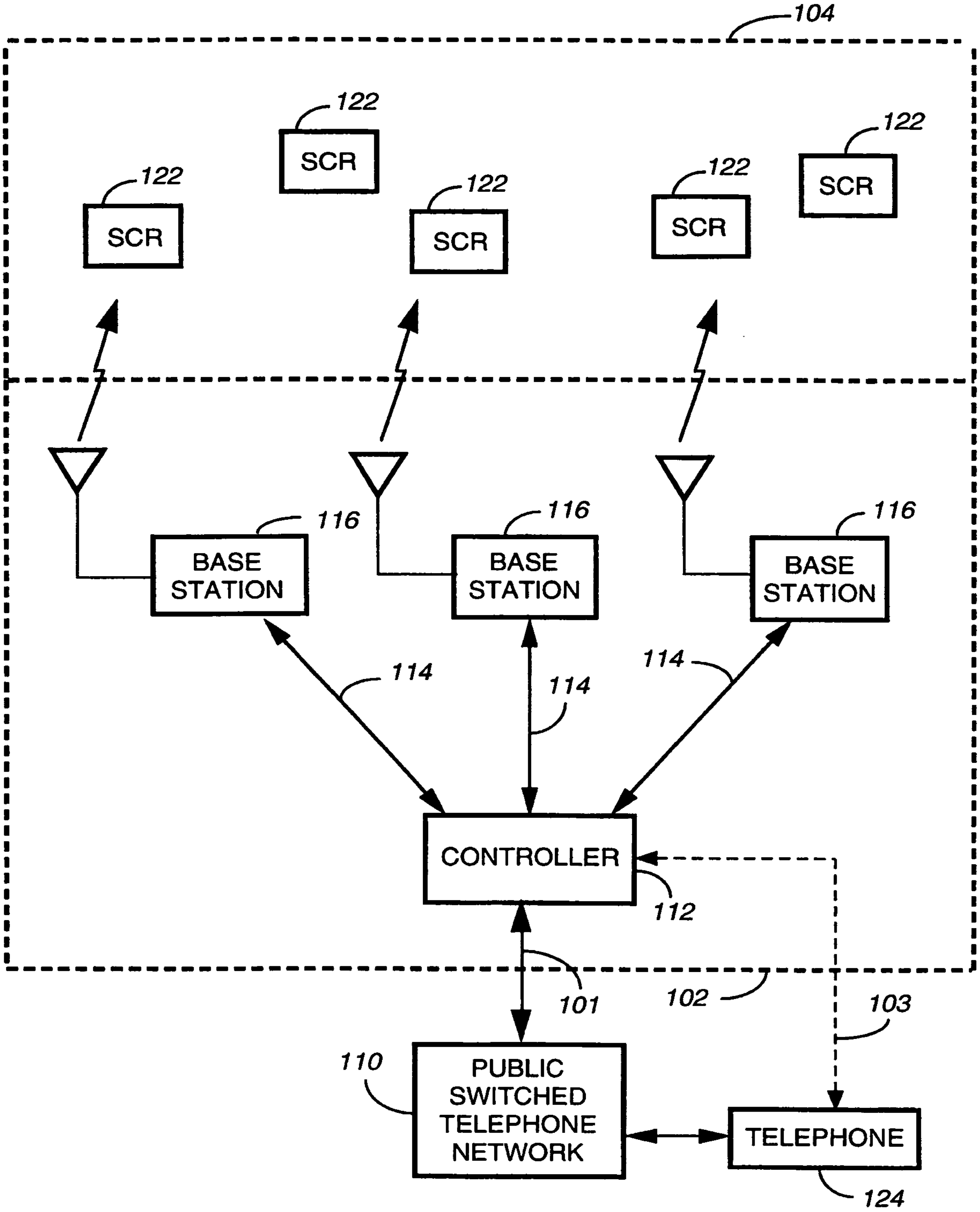
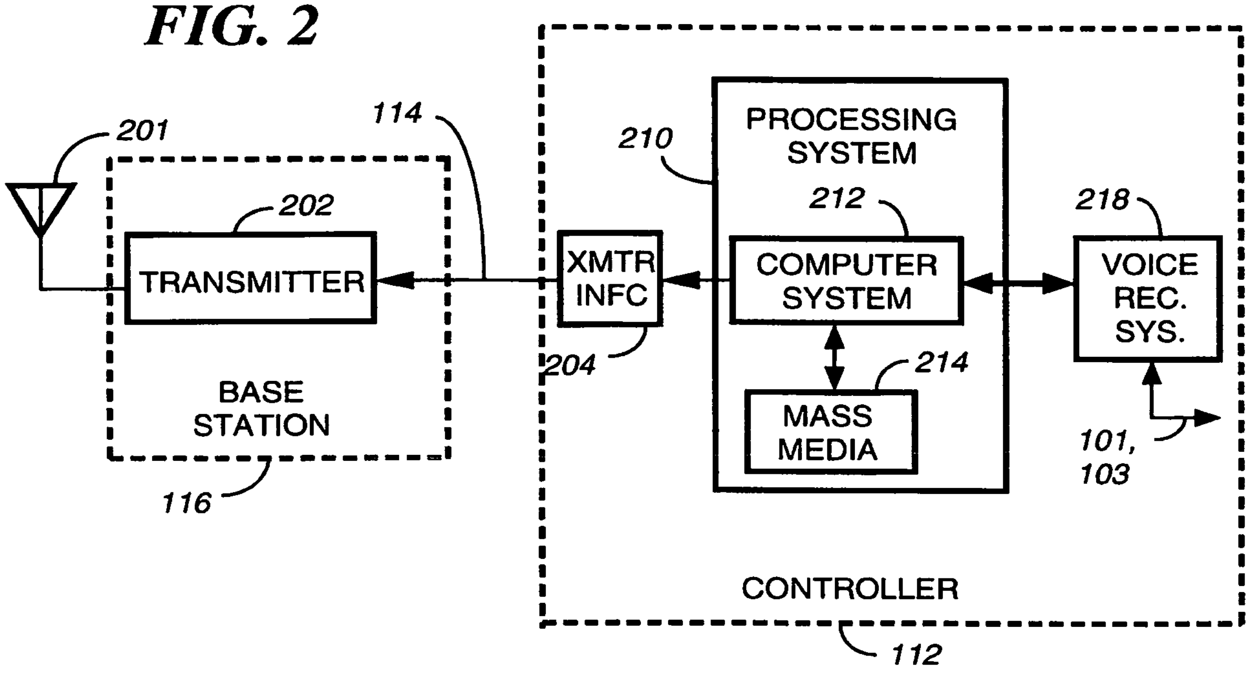

104
122
SCR
122
SCR
122
SCR
122
SCR
122
SCR
116
BASE STATION
116
BASE STATION
116
BASE STATION
114
114
114
CONTROLLER
112
101
102
103
110
PUBLIC SWITCHED TELEPHONE NETWORK
TELEPHONE
124

FIG. 1

FIG. 2



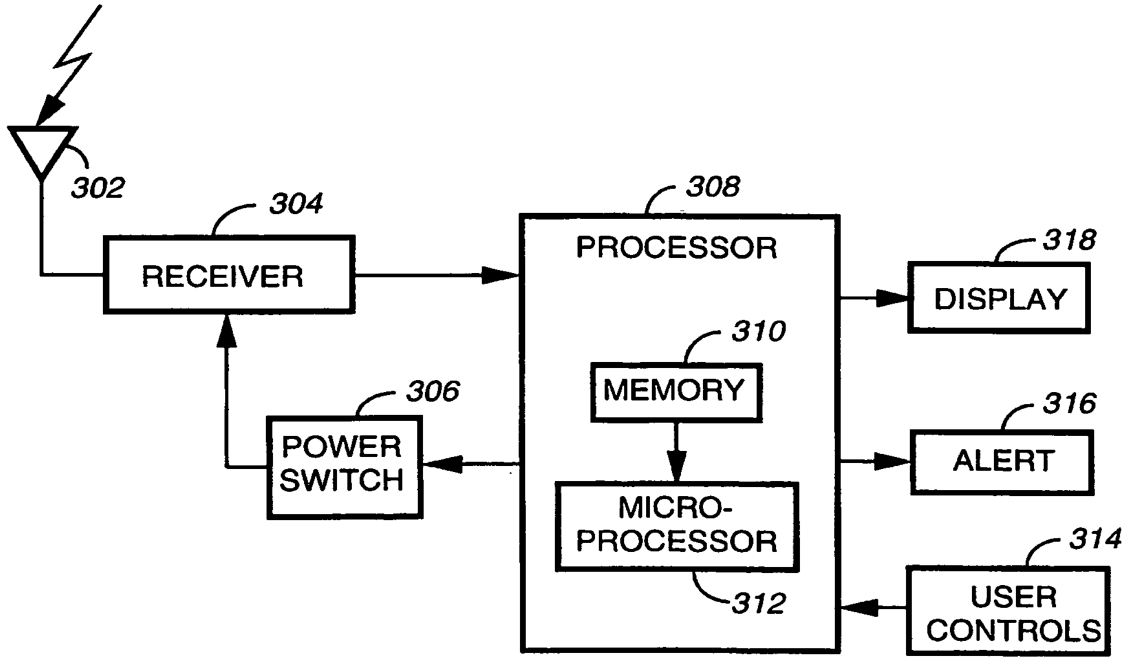


FIG. 3 122

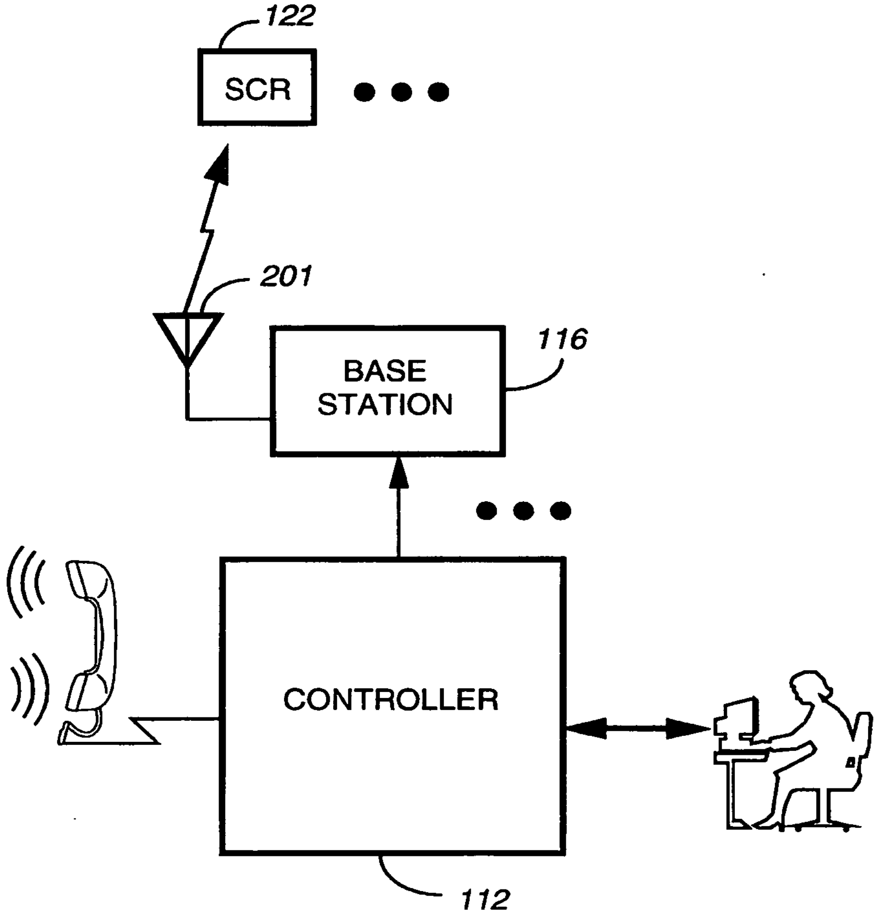
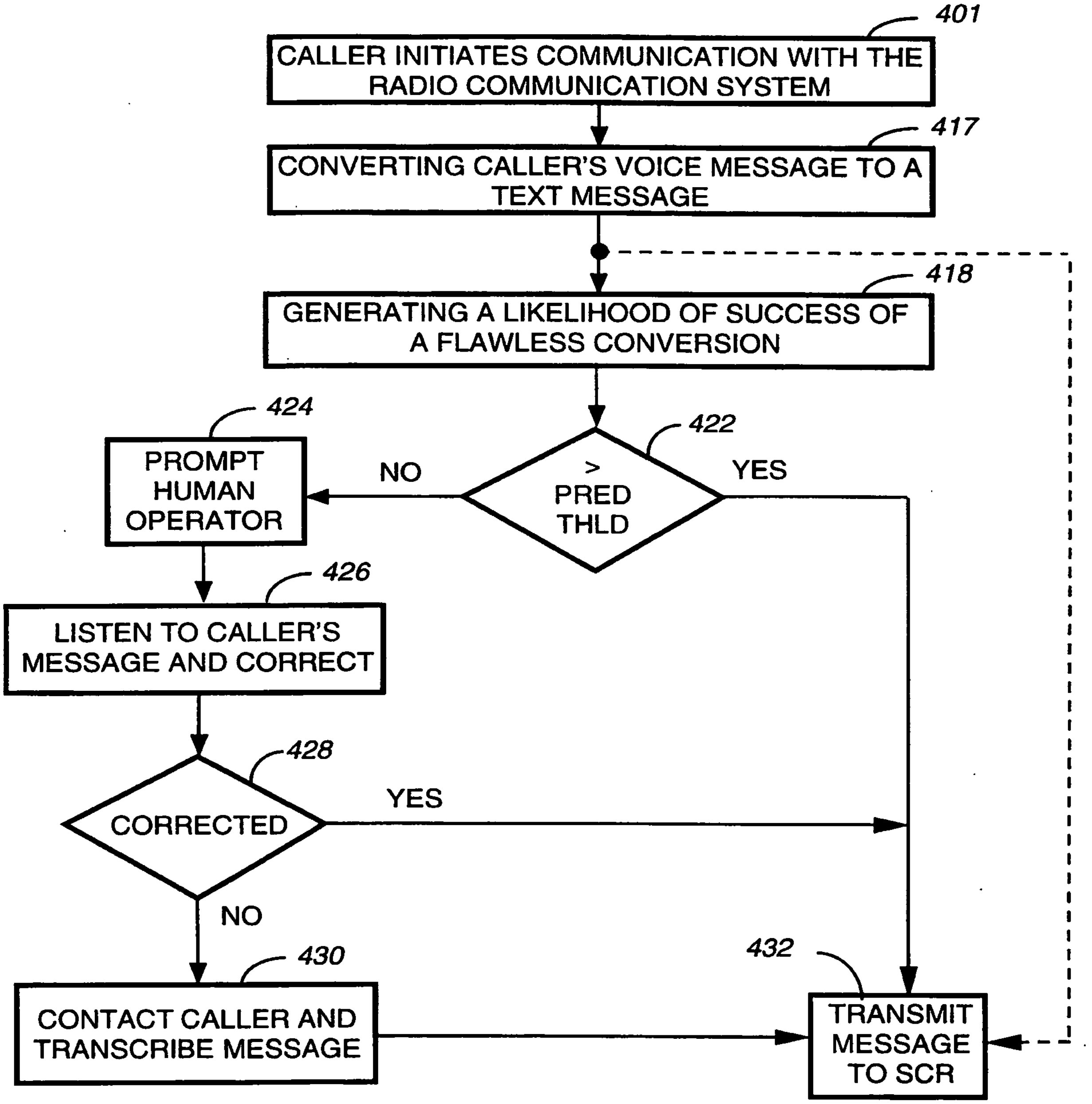


FIG. 4


401
CALLER INITIATES COMMUNICATION WITH THE RADIO COMMUNICATION SYSTEM
417
CONVERTING CALLER'S VOICE MESSAGE TO A TEXT MESSAGE
418
GENERATING A LIKELIHOOD OF SUCCESS OF A FLAWLESS CONVERSION
422
> PRED THLD
NO
YES
424
PROMPT HUMAN OPERATOR
426
LISTEN TO CALLER'S MESSAGE AND CORRECT
428
CORRECTED
YES
NO
430
CONTACT CALLER AND TRANSCRIBE MESSAGE
432
TRANSMIT MESSAGE TO SCR

400

FIG. 5

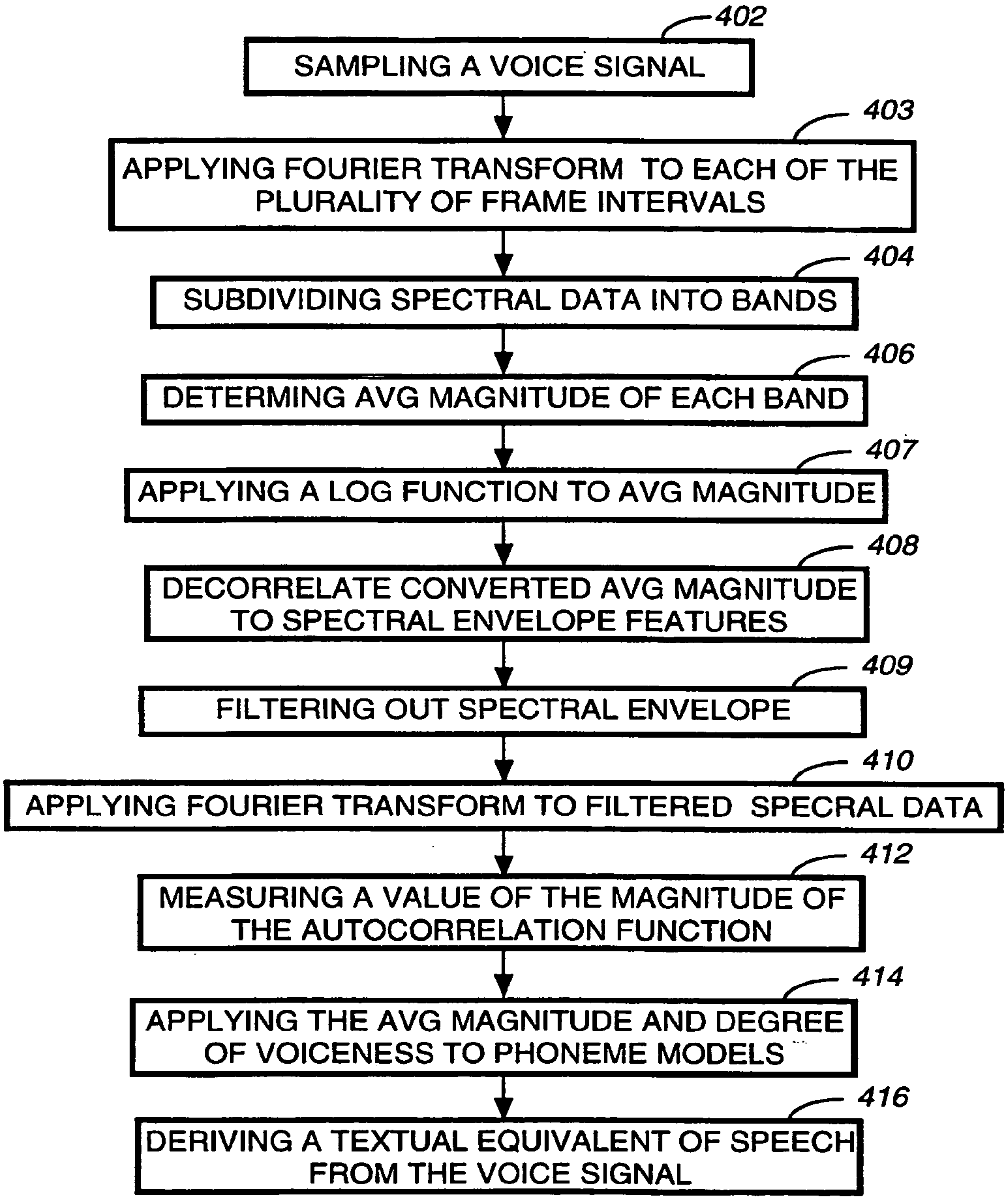

402
SAMPLING A VOICE SIGNAL
403
APPLYING FOURIER TRANSFORM TO EACH OF THE PLURALITY OF FRAME INTERVALS
404
SUBDIVIDING SPECTRAL DATA INTO BANDS
406
DETERMING AVG MAGNITUDE OF EACH BAND
407
APPLYING A LOG FUNCTION TO AVG MAGNITUDE
408
DECORRELATE CONVERTED AVG MAGNITUDE TO SPECTRAL ENVELOPE FEATURES
409
FILTERING OUT SPECTRAL ENVELOPE
410
APPLYING FOURIER TRANSFORM TO FILTERED SPECRAL DATA
412
MEASURING A VALUE OF THE MAGNITUDE OF THE AUTOCORRELATION FUNCTION
414
APPLYING THE AVG MAGNITUDE AND DEGREE OF VOICENESS TO PHONEME MODELS
416
DERIVING A TEXTUAL EQUIVALENT OF SPEECH FROM THE VOICE SIGNAL
417

FIG. 6

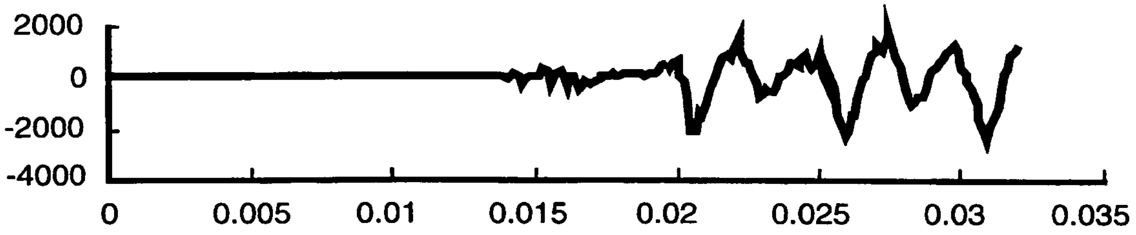


*FIG. 7*

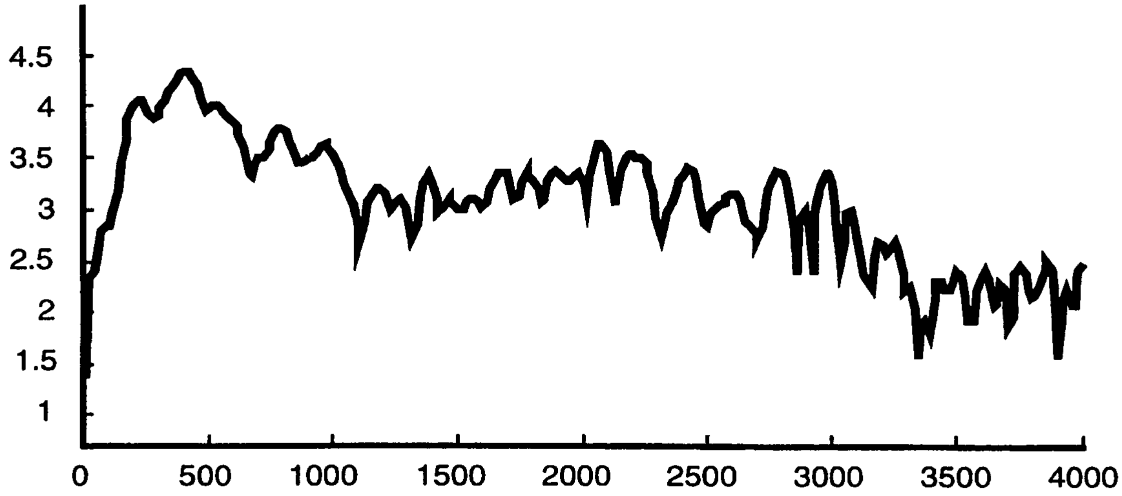


*FIG. 8*

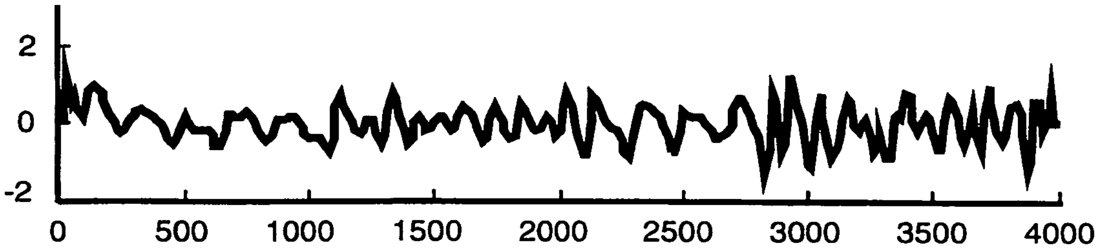


*FIG. 9*

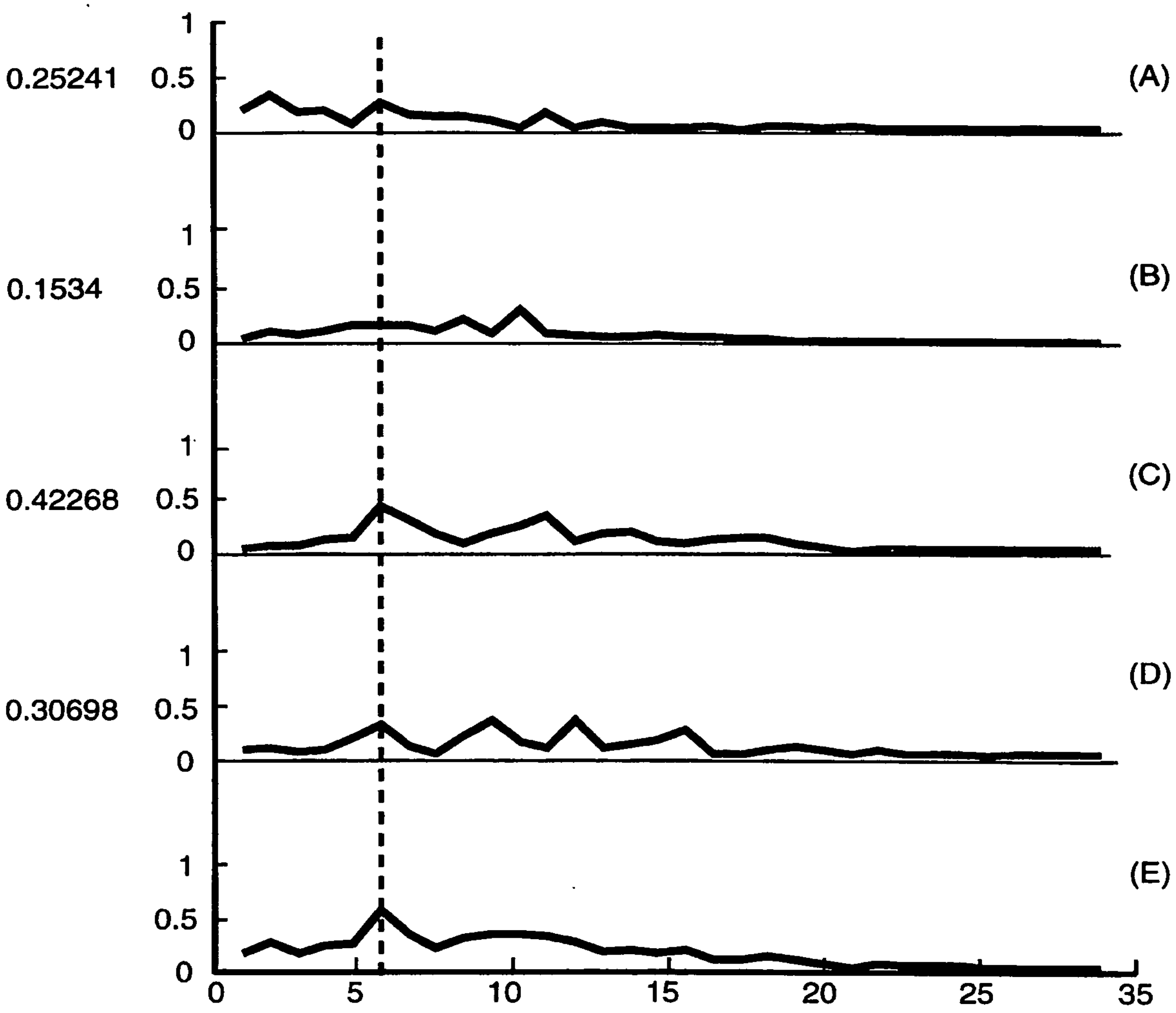

1
0.25241
0.5
0
(A)
1
0.1534
0.5
0
(B)
1
0.42268
0.5
0
(C)
1
0.30698
0.5
0
(D)
1
0.5
0
(E)
0
5
10
15
20
25
35

FIG. 10

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/06600

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G10L 5/06; H04M 1/64
US CL :704/235; 455/563
According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/235, 240; 455/563; 379/88.14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS; DIALOG

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 98/05154 A1 (TELIA AB) 05 February 1998 (05.02.98), claims, figure 1. | 10,21 |
| X,P | US 5,870,454 A (DAHLEN) 09 February 1999 (09.02.99), columns 1-3, 7, 8. | 10,21 |
| Y | US 5,396,542 A (ALGER ET AL) 07 March 1995 (07.03.95), column 2. | 1,2,10,21,18,19 |
| Y | US 5,712,901 A (MEERMANS) 27 January 1998 (27.01.98), columns 2, 4, 6. | 2,19 |

☒ Further documents are listed in the continuation of Box C. □ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

10 JUNE 1999

Date of mailing of the international search report

12 JUL 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
DAVID R. HUDSPETH Joni Hill
Telephone No. (703) 308-0956

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/06600

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | US 5,390,278 A (GUPTA ET AL) 14 February 1995 (14.02.95), column 13. | 1,10,18,21 |
| Y | US 5,719,996 A (CHANG ET AL) 17 February 1998 (17.02.98), columns 2, 3, 6. | 1,3,5,7-11, 13,15-18,2 0-22 |
| Y | US 4,058,676 A (WILKES ET AL) 15 November 1977 (15.11.77), columns 2, 8-11, 13-15, 18-23. | 3,5-9,11,13-17,20,22 |
| Y | US 5,828,993 A (KAWAUCHI) 27 October 1998 (27.10.98), columns 1,2 | 3,8,11,16,20,22 |
| Y | US 5,600,703 A (DANG ET AL) 04 February 1997 (04.02.97), columns 1, 5-9, 12. | 1,10,18,21 |
| Y | US 5,566,272 A (BREMS ET AL) 15 October 1996 (15.10.96), columns 1, 3-6. | 1,2,10,18,21 |